SIOG 231G Electrical Methods in Geophysics

Lecture 9a: Geomagnetic Depth Sounding

Geomagnetic Depth Sounding

If we want to look deep into Earth using electrical techniques, to study the properties of the deeper mantle, then the MT method is not the ideal tool. Firstly, you need very long time series of the EM fields, spanning perhaps years, and while magnetic observatories have been collecting records for more than a hundred years, there are few observatories collecting electric field data. Secondly, if we examine the MT relationship:

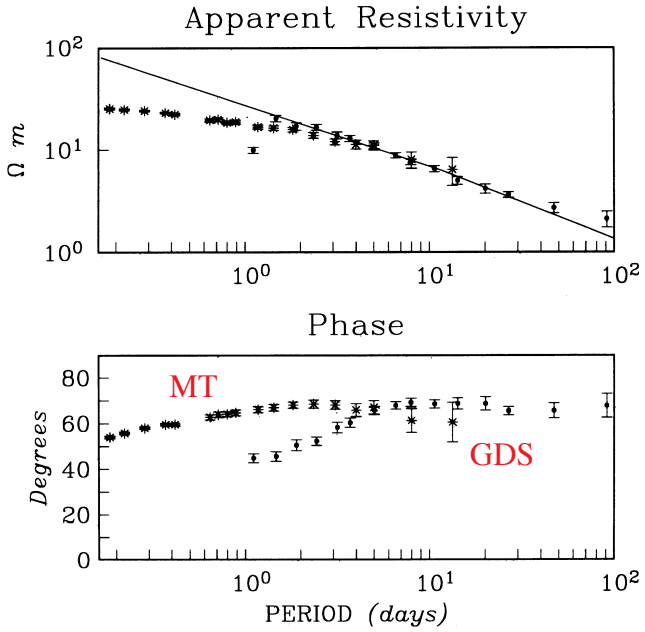
$$\rho_a = \frac{T\mu_o}{2\pi}\left|\frac{E}{B}\right|^2$$

(where we have replaced frequency with period, T), we observe that for a given magnetic field and resistivity the amplitude of the induced electric field decreases with period:

$$E = B\sqrt{\frac{2\pi\rho_a}{T\mu_o}}$$

which is exacerbated by a falling ρ_a with depth and period. At long periods the assumption that the source field morphology is uniform over scales comparable to the depth of penetration in Earth breaks down, particularly for the daily variation and harmonics. Finally, there will be an ambiguity in the absolute value of mantle conductivity associated with unknown static/galvanic effects in the electrical data caused by lateral conductivity contrasts near Earth's suface.

We can derive other types of electromagnetic response functions, similar in nature to the magnetotelluric response, using only the three components of the magnetic field recorded by geomagnetic observatories (geomagnetic depth sounding, or GDS, responses). Much of this work is still based on techniques developed by Banks (1969). The figure below shows a combination of MT and GDS response functions from Egbert and Booker (1992). Magnetotelluric data are available to a maximum of 10 days period (about 10^6 seconds), while GDS data can extend this to periods of many months.



Using Observatory Records

We can again write the observed field as the gradient of a scalar potential Φ expressed as a spherical harmonic expansion of associated Legendre polynomials P_l^m, with Schmidt quasi-normalized spherical harmonic coefficients representing the internal i_l^m and external e_l^m magnetic fields:

$$\Phi(r,\theta,\phi) \;=\; a_o \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \left\{ i_l^m \left(\frac{a_o}{r} \right)^{l+1} \;+\; e_l^m \left(\frac{r}{a_o} \right)^{l} \right\} P_l^m(\cos\theta) e^{im\phi} \quad .$$

Previously, we carried out the spherical harmonic expansion in geocentric coordinates, but here we will use geomagnetic coordinates in order to minimize non-axisymmetric terms. Geomagnetic coordinates are defined by the axis of the best-fitting dipole, and so necessarily will vary over time as secular variation occurs. Note that by using the exponential term the Gauss coefficients i_l^m and e_l^m are implicitly complex in this equation. Now θ is *geomagnetic* co-latitude, but a_o stays the radius of Earth and r remains the radius of observation.

Let us assume that the non axisymmetric terms are small enough to be neglected. Then we lose the terms in m and

$$\Phi = a_o \sum_{l=1}^{\infty} \left\{ i_l \left(\frac{a_o}{r} \right)^{l+1} + e_l \left(\frac{r}{a_o} \right)^{l} \right\} P_l^0(cos\theta) \quad .$$

Notice that by making $m = 0$ the exponential term $e^{im\phi}$ goes away and the Gauss coefficients become real. The i_l and e_l and can be functions of time or frequency when we consider time variations in the external field. If we start with magnetic observatory measurements, it is usual to work in the frequency domain, but note that conversion to frequency gets us back into the complex plane in order to capture both amplitude and phase.

We can define a geomagnetic response for a radially symmetric Earth as the ratio of induced (internal) to external fields:

$$Q_l(\omega) = \frac{i_l(\omega)}{e_l(\omega)}$$

where the frequency domain is made explicit, and will be implicit in the following. This definition makes sense because for a given external field variation, the bigger the electrical conductivity, the bigger the induced field. Also, the lower the frequency, the larger the skin depth and the deeper the depth of the induced fields.

In certain circumstances, such as satellite observations or using a global array of observatories, we have enough data to fit the e_l and i_l directly. However, much of the time we consider just the horizontal and vertical components of **B** as recorded by a single observatory. (Since we are working in geomagnetic coordinates and have ignored the non-axisymmetric components, the horizontal component always points north or south.) We can obtain expressions for the horizontal component H of **B** and the vertical component Z of **B** at the surface of Earth ($r = a_o$) from the appropriate partial derivatives of Φ:

$$H = \left(-\frac{1}{r} \frac{\partial \Phi}{\partial \theta} \right)_{r=a_o} = \left(\frac{a_o}{r} \sum_{l}^{\infty} \left\{ i_l \left(\frac{a_o}{r} \right)^{l+1} + e_l \left(\frac{r}{a_o} \right)^{l} \right\} \frac{\partial P_l^0(\cos\theta)}{\partial \theta} \right)_{r=a_o}$$

$$= \sum_{l} A_{H,l} \frac{\partial P_l^0(\cos\theta)}{\partial \theta}$$

and

$$Z = \left(-\frac{\partial \Phi}{\partial r} \right)_{r=a_o} = \left(a_o \sum_{l}^{\infty} \left\{ -i_l a_o^{l+1} (l+1) r^{-l-2} + e_l a_o^{-l} (l) r^{l-1} \right\} P_l^0(cos\theta) \right)_{r=a_o}$$

$$= \sum_{l} A_{Z,l} P_l^0(\cos\theta)$$

where we have defined new expansion coefficients

$$A_{H,l} = i_l + e_l$$

$$A_{Z,l} = le_l - (l+1)i_l \quad .$$

Two equations in two unknowns, so we can solve for i_l and e_l:

$$i_l = \frac{lA_{H,l} - A_{Z,l}}{2l+1} \qquad e_l = \frac{A_{Z,l} + (l+1)A_{H,l}}{2l+1}$$

and our electromagnetic response Q is related to the A's by

$$Q_l = \frac{i_l}{e_l} = \frac{l - A_{Z,l}/A_{H,l}}{l+1+A_{Z,l}/A_{H,l}} = \frac{l - W_l}{l+1+W_l}$$

where we have introduced the ratio of A's as a new electromagnetic response

$$W_l = \frac{A_{Z,l}}{A_{H,l}} \quad .$$

It turns out that the inductive scale length or c-response is related to W_l by

$$c_l = \frac{a_o W_l}{l(l+1)} \quad .$$

By looking at the geometry of simultaneous observatory records as a function of geomagnetic colatitude, Roger Banks showed that except for the daily and annual variations, the magnetic field was dominantly of P_1^0 geometry as a result of the nature of the ring current. Thus, the above analysis could be simplified some more by assuming the simple degree-one field geometry. In this case W becomes simply the ratio of vertical to horizontal fields with a co-latitudinal trigonometric term:

$$W = \frac{A_{Z,1}}{A_{H,1}} = \frac{Z/P_1^0(\cos\theta)}{H/\frac{\partial}{\partial\theta}P_1^0(\cos\theta)}$$

and recalling that $P_1^0(\cos\theta) = \cos\theta$,

$$W = \frac{Z\sin\theta}{H\cos\theta} = \frac{Z(\omega)}{H(\omega)}\tan\theta$$

remembering that $Z(\omega)/H(\omega)$ is complex and a function of frequency. (This method blows up at $\theta = [0, 90, 180]$. Why?) The c-response becomes simply $c = a_o W/2$ and the magnetotelluric apparent resistivity and phase are

$$\rho_a = \omega\mu_o|c|^2 \qquad \phi = \arg(c) \quad .$$

The workflow for geomagnetic depth sounding is to take H and Z records from a single observatory and Fourier transform them to get the field ratio as a function of frequency. One could take the individual records, transform them, and then take the ratio of the Fourier components, but since there is noise in both the records this isn't great. A better way is to take a cross-spectrum which pulls out the coherent signal in both records. Since some observatories have been recording for over a hundred years, this allows response functions to be computed out to periods of at least 6 months.

Of course, by using a number of observatories distributed across colatitudes, one could fit the P_1^0 field geometry rather than assuming it *a priori* . This is rarely done for a number of reasons. First, not all observatories are created equal, in terms of noise in the records and the length of time they have been operating. However, the bigger reason is that even though the above analysis implicitly assumes the electrical conductivity is radially symmetric, there will be variations in conductivity beneath various observatories, which can be identified by variations in the response functions.

Magnetic Satellite Studies

Just as magnetic satellites have provided additional data for main field modeling, they provide another way to estimate GDS responses. They have the advantage that they provide continuous data from pole to pole, and sample the entire Earth. However, because satellites move through the static parts of Earth's magnetic field, folding spatial variations into temporal variations within the satellite frame of reference, considerable effort is required to remove

- The main (core) field, and its secular variation.
- The crustal field due to remanent and induced magnetization.
- Ionospheric currents (daily variation)
- Field aligned and meridional currents, and seasonal variations.
- Equatorial electrojet.
- Coupling and induction of the above.

One tool that attempts to model all these phenomena simultaneously is the Comprehensive Model (Sabaka *et al.*, 2004). Even so, it is difficult to remove the effects of field aligned currents in the auroral zones, and so data are typically processed between $\pm 50^\circ$ geomagnetic latitude, and residual effects of the daily variation are usually avoided by using only night-time data.

Because we have continuous data as a function of geomagnetic colatitude, rather than assume a P_1^0 geometry at a single position as we did for observatory data, we can compute the best fitting P_1^0 function directly from the magnetic field data. We do this for each pass to obtain a time series of internal and external Gauss coefficients. Only after that do we convert to the frequency domain.

As above, we keep only the P_1^0 contribution and use r, θ, ϕ in geomagnetic coordinates, and can then write the magnetic potential as

$$\Phi_1^0(r,\theta,\phi) \;=\; a_o\left\{ i_1^0(t)\left(\frac{a_o}{r}\right)^2 \;+\; e_1^0(t)\left(\frac{r}{a_o}\right)\right\} P_1^0\big(\cos\theta\big)$$

where i_1^0 and e_1^0 are real and we have made the function of time t explicit. The magnetic field $\mathbf{B}$ is derived from the negative of the gradient in the usual manner

$$\mathbf{B}(r,\theta,\phi) = -\nabla\Phi_1^0(r,\theta,\phi)$$

recalling that the gradient operator in spherical coordinates looks like

$$\nabla\Phi = \left(\frac{\partial\Phi}{\partial r}, \frac{1}{r}\frac{\partial\Phi}{\partial\theta}, \frac{1}{r\sin\theta}\frac{\partial\Phi}{\partial\phi}\right)$$

and that $P_1^0(\cos\theta) = \cos\theta$, we can write the components B_r, B_θ, B_ϕ of our spherical coordinate system:

$$B_r = -\frac{\partial\Phi}{\partial r} = \left[-e_1^0 + 2i_1^0\left(\frac{a}{r}\right)^3\right]\cos(\theta)$$

$$B_\theta = -\frac{1}{r}\frac{\partial\Phi}{\partial\theta} = \left[e_1^0 + i_1^0\left(\frac{a}{r}\right)^3\right]\sin(\theta)$$

$$B_\phi = 0 \quad .$$

The azimuthal (ϕ) component is zero because we have neglected all non-zonal components, and as for the observatory records, when you do this the horizontal component is purely in the θ direction (north-south). Note that the altitude of satellites varies along the orbit, so we need to keep the r dependence, unlike for observatory studies. The above equations may be expressed in matrix form as:

$$\begin{bmatrix} -\cos(\theta) & 2(a/r)^3\cos(\theta) \\ \sin(\theta) & (a/r)^3\sin(\theta) \end{bmatrix} \begin{bmatrix} e_1^0 \\ i_1^0 \end{bmatrix} = \begin{bmatrix} B_r \\ B_\theta \end{bmatrix} \quad .$$

For an analysis using a single satellite, these equations can be fit to data from each satellite pass from north to south (or vice versa) (e.g. Figure 4) having different values of altitude (r) and geomagnetic colatitude (θ), as was done by Constable and Constable (2004b), providing estimates of $i_1^0(t)$ and $e_1^0(t)$ at about 100 minute intervals. Residuals from the fitting were about 5–10 nT.

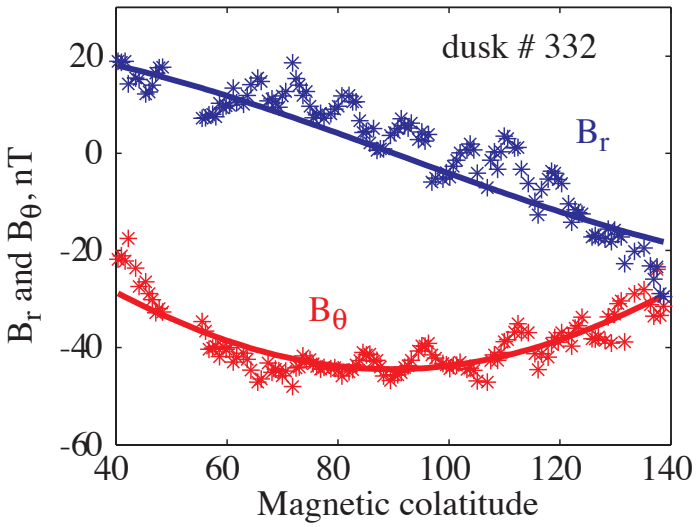


Figure 4. Fits (solid lines) to data from a single 90-minute satellite pass (symbols) . From these fits we obtain two real numbers, i_1^0 and e_1^0 for a time t taken to be the mean time of the pass.

The i_1^0 and e_1^0 estimates for single passes can be merged to form a time series for the entire satellite mission (Figure 5). It can be seen that the magnitude of both the internal and external fields are larger for the dusk passes than the dawn passes, clearly indicating that the ring current is not symmetric about Earth, as often assumed by the induction community, a feature well understood by the space physics community.

Response functions may be obtained from the $i_1^0(t)$ and $e_1^0(t)$ time series in the usual way by transformation into the frequency domain to derive a complex geomagnetic response function of frequency:

$$\mathrm{Q}_1^0(\omega) = \frac{i_1^0(\omega)}{e_1^0(\omega)}$$

from which we get admittance by

$$c(\omega) = a_o \frac{l - (l+1)Q_l}{l(l+1)(1+Q_l)} = a_o \frac{1 - 2Q_1}{2 + 2Q_1}$$

where a_o is still the radius of Earth and l is the order of the spherical harmonic expansion, which we have set to 1 in this case.

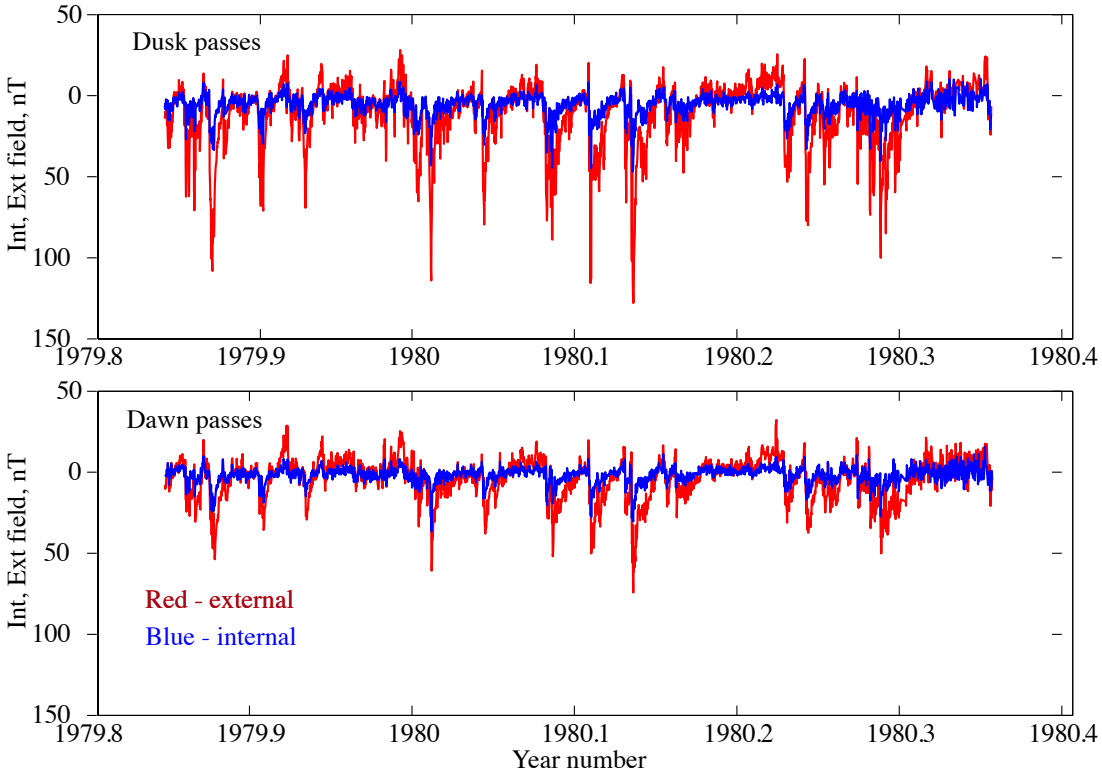


Figure 5 Time series of $i_1^0(t)$ and $e_1^0(t)$ for the entire MAGSAT mission, separated into dawn and dusk passes.

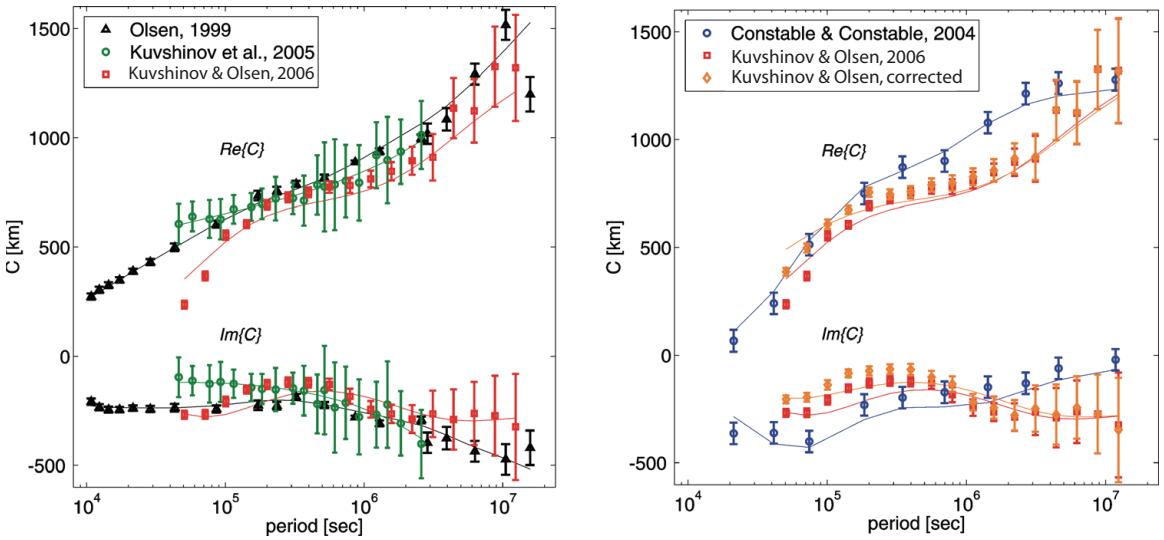


Figure 6. Long period geomagnetic response functions computed from magnetic satellite records, after Kuvshinov and Olsen (2006). The Olsen (1999) response is a traditional GDS response made using data from European observatories. Kuvshinov and Olsen attribute the difference in the long period imaginary response between their data and those of Constable and Constable (2004) as an artifact of Constables' time series analysis. Their correction removes the 3D effects of the oceans on the 1D response functions.

Lecture 9b: Linear and Nonlinear Least Squares Fitting

Introduction

In geophysics we fit models to data all the time. The model could be a straight line (a classic problem found on most programmable calculators in the past), a more complex but smooth trend you want to remove from a series of data, a large but linear system (think lots of Gauss coefficients to fit the main magnetic field), a small number of parameters describing a nonlinear problem (particle density and mobility describing conduction in a semiconductor as a function of temperature – we will see this later), or a large number of parameters describing a nonlinear problem (a global 3D electrical conductivity model or even a local 2D conductivity model). We will see all these things in this course. The time-honored, and often the optimal way, to do this is by least-squares fitting.

Least squares as a measure

We start with a set of M observed data $\mathbf{d}$:

$$\mathbf{d} = (d_1, d_2, d_3, ..., d_M)^T \quad .$$

These data will have errors, or uncertainty, associated with them:

$$\sigma = (\sigma_1, \sigma_2, ..., \sigma_M)^T$$

which may be estimated during the generation of the data set, an educated guess, or an assumption that all the errors are the same (in which case the size of the constant error won't matter). In practice, getting good error estimates can be as hard as getting the actual data.

We have some mathematical description f which captures the physics of how some model $\mathbf{m}$, which describes our world, is related to the data:

$$\hat{\mathbf{d}} = f(\mathbf{x}, \mathbf{m})$$

where $\hat{\mathbf{d}}$ are the predicted response (i.e. estimates of the data given the model $\mathbf{m}$):

$$\hat{\mathbf{d}} = (\hat{d}_1, \hat{d}_2, \hat{d}_3, ..., \hat{d}_M)^T$$

and $\mathbf{x}$ are independent data variables which describe the data (for GDS, frequencies or periods, for magnetic field measurements, location on the globe) associated with the predicted responses. Let's assume we only need one per datum, although in practice several might be required (e.g. latitude, longitude, time, and an index of field component if you are fitting a secular variation magnetic field model):

$$\mathbf{x} = (x_1, x_2, x_3, ..., x_M) \quad .$$

Note that in the real world our model could be infinite dimensional, even for simple looking models. For example, if we wanted to model electrical conductivity as a function of depth in Earth, there are an infinite number of possible depths. For this reason $f(\mathbf{m})$ is called a forward functional, because it maps a potentially infinite dimensional model into a single data point at a time. In practice, however, we discretize our model into a finite number of N parameters:

$$\mathbf{m} = (m_1, m_2, ..., m_N)^T \quad .$$

The number of parameters could be very large in practice, especially if we want to approximate an infinite dimensional problem, and we will get to how you handle such cases, but for now let's assume that $N < M$ (the size of the model is smaller than the number of data).

One might think that we'd like a scheme for finding a model $\mathbf{m}_*$ that exactly fits the observed data $\mathbf{d}$ such that

$$\mathbf{d} = f(\mathbf{x}, \mathbf{m}_*) \quad .$$

In practice, it is unlikely that any model exists that will fit the data, which always have noise, perfectly, and it is often true that your parameterization does not capture the complexities of the real world that generated the data. What we do instead is define a measure of how well a model **m** fits the data, and find some way to reduce, or even minimize, this misfit. For practical and theoretical reasons, the sum-squared misfit is a favored measure:

$$\chi^2 = \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \left[d_i - f(x_i, \mathbf{m}) \right]^2$$

which equivalently may be written as

$$\chi^2 = ||\mathbf{W}(\mathbf{d} - \hat{\mathbf{d}})||^2 = ||\mathbf{W}\mathbf{d} - \mathbf{W}f(\mathbf{m})||^2$$

where **W** is a diagonal matrix of reciprocal data errors

$$\mathbf{W} = \text{diag}(1/\sigma_1, 1/\sigma_2, ...1/\sigma_M) \quad .$$

The least squares (LS) approach attempts to minimize χ^2 with respect to all the model parameters simultaneously. If the data errors σ_i are Gaussian (normally distributed) and independent, then least squares provides a maximum liklihood and unbiased estimate of $\mathbf{m}_*$, and χ^2 is Chi-squared distributed with $M - N$ degrees of freedom. It turns out that having independent, normally distributed data is a fairly high bar, but the method is fairy tolerant to deviations from this ideal.

Linear problems

For linear problems, our forward functional $f(\mathbf{m})$ can be written as

$$\hat{\mathbf{d}} = \mathbf{F}\mathbf{m}$$

or

$$\mathbf{d} = \mathbf{F}\mathbf{m} + \mathbf{r}$$

where **F** is an $M \times N$ matrix, often called the design matrix, which linearly maps the model into the data, and **r** is an M-column vector of residuals. If we write

$$\mathbf{r} = \mathbf{d} - \mathbf{F}\mathbf{m}$$

the sum of squares misfit is $\mathbf{r}^T\mathbf{r}$, which we can minimize by differentiating and setting to zero, to get the least squares solution:

$$\mathbf{m}_* = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{d}$$

(don't worry if this went fast - you will see the derivation, and a lot more, in SIOG 223 and SIOG 230). These are called the normal equations.

What does **F** look like? Recall our P_1^0 fitting of satellite data:

$$\begin{bmatrix} B_r \\ \\ B_\theta \end{bmatrix} = \begin{bmatrix} -\cos(\theta) & 2(a/r)^3\cos(\theta) \\ \\ \sin(\theta) & (a/r)^3\sin(\theta) \end{bmatrix} \begin{bmatrix} e_1^0 \\ \\ i_1^0 \end{bmatrix} \quad .$$

This maps a single pair of data B_r and B_θ to a pair of Gauss coefficients. But looking at Figure 4 we have lots of data (123 for each component to be precise) at lots of magnetic colatitudes. So what we really have is

$$\begin{bmatrix} B_r(\theta_1, r_1) \\ B_\theta(\theta_1, r_1) \\ B_r(\theta_2, r_2) \\ B_\theta(\theta_2, r_2) \\ \cdots \\ B_r(\theta_{123}, r_{123}) \\ B_\theta(\theta_{123}, r_{123}) \end{bmatrix} = \begin{bmatrix} -\cos(\theta_1) & 2(a/r_1)^3\cos(\theta_1) \\ \sin(\theta_1) & (a/r_1)^3\sin(\theta_1) \\ -\cos(\theta_2) & 2(a/r_2)^3\cos(\theta_2) \\ \sin(\theta_2) & (a/r_2)^3\sin(\theta_2) \\ \cdots & \cdots \\ -\cos(\theta_{123}) & 2(a/r_{123})^3\cos(\theta_{123}) \\ \sin(\theta_{123}) & (a/r_{123})^3\sin(\theta_{123}) \end{bmatrix} \begin{bmatrix} e_1^0 \\ i_1^0 \end{bmatrix} \quad .$$

Here the vector of magnetic colatitudes $[\theta_1, \theta_2, ...\theta_{123}]$ are equivalent to the independent data variables **x** described above.

Nonlinear problems

What happens if your forward solution is not linear? First, see if you can linearize it. We will see that a good model for the electrical conductivity of minerals is the Arrhenius relation

$$\sigma(T) = \sigma_o e^{-A/kT}$$

where T is absolute temperature, σ_o is a constant, A is activation energy, and k is Boltzmann's constant. This can be linearized as

$$\log(\sigma(T)) = \log(\sigma_o) - (A/k)(1/T) \quad .$$

However, if this is not possible, most practical approaches to the least squares problem usually involve linearizing the forward solution f around an initial model guess $\mathbf{m}_0$

$$f(\mathbf{m}_0 + \Delta\mathbf{m}) = f(\mathbf{m}_0) + \mathbf{J}\Delta\mathbf{m} + O(\Delta\mathbf{m}^2)$$

where **J** is a matrix of partial derivatives of data with respect to the model parameters

$$J_{ij} = \frac{\partial f(x_i, \mathbf{m}_0)}{\partial m_j}$$

(often called the Jacobian matrix) and

$$\Delta\mathbf{m} = (\delta m_1, \delta m_2, ..., \delta m_N)$$

is a model parameter perturbation about $\mathbf{m}_0$. Now our expression for χ^2 is

$$\chi^2 \approx ||\mathbf{W}(\mathbf{d} - f(\mathbf{m}_0)) + \mathbf{W}\mathbf{J}\Delta\mathbf{m}||^2$$

where the approximation is a recognition that we have dropped the higher order terms. We will proceed on the assumption that this linear approximation is a good one. (More on that later.) We can minimize χ^2 in the usual way by setting the derivatives of χ^2 with respect to $\Delta\mathbf{m}$ equal to zero:

$$\nabla\chi^2 = -2(\mathbf{W}\mathbf{J})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0})) - \mathbf{W}\mathbf{J}\Delta\mathbf{m}] = 0$$

and solving for $\Delta\mathbf{m}$

$$\Delta\mathbf{m} = [(\mathbf{W}\mathbf{J})^T\mathbf{W}\mathbf{J}]^{-1}(\mathbf{W}\mathbf{J})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))]$$

which may be equivalently written as N simultaneous equations:

$$\beta = \alpha\Delta\mathbf{m}$$

where

$$\beta = (\mathbf{W}\mathbf{J})^T\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))$$

$$\alpha = (\mathbf{W}\mathbf{J})^T\mathbf{W}\mathbf{J} \quad .$$

These are equivalent to our normal equations for the linear problem, and the matrix α is sometimes called the curvature matrix. This system can be solved for $\Delta\mathbf{m}$ by inverting α numerically. If the forward model f is truly linear, then the model $\mathbf{m}_* = \mathbf{m}_0 + \Delta\mathbf{m}$ is the least squares solution.

For non-linear problems a second model $\mathbf{m}_1 = \mathbf{m}_0 + \Delta\mathbf{m}$ is found and the process repeated in the hope that the process converges to a solution. Note that because **J** depends on **m**, **J** will need to be re-computed every iteration. This is the general form of the Gauss-Newton approach to model fitting.

Near the least squares solution, the Gauss-Newton method will work, but any significant nonlinearity will result in likely failure unless you start linearly "close" to a solution. Long, thin, "valleys" in the χ^2 hyper-surface are common and produce search directions which diverge radically from the direction of the minimum.

Another approach which does not depend on how large the second-order terms in the expansion of f are is based on the expansion of χ^2, rather than f:

$$\chi^2(\mathbf{m}_0 + \Delta\mathbf{m}) = \chi^2(\mathbf{m}_0) + \Delta\mathbf{m}^T \nabla\chi^2(\mathbf{m}_0) + O(\Delta\mathbf{m}^2)$$

into which we can substitute our expression (19) for $\nabla\chi^2$ (setting $\Delta\mathbf{m} = 0$) and drop the high order terms

$$\chi^2(\mathbf{m}_0 + \Delta\mathbf{m}) = \chi^2(\mathbf{m}_0) - 2\Delta\mathbf{m}^T(\mathbf{WJ})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))] \; .$$

If we choose $\Delta\mathbf{m} = \mu(\mathbf{WJ})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))]$ for some scalar μ then we get

$$\chi^2(\mathbf{m}_0 + \Delta\mathbf{m}) = \chi^2(\mathbf{m}_0) - \mu||(\mathbf{WJ})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))]||^2$$

and there will always be a μ which keeps reducing χ^2 (not all μ – at some point the higher order terms will take over). Solutions of the form

$$\Delta\mathbf{m} = \mu(\mathbf{WJ})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))]$$

go down the slope of the χ^2 hyper-surface, and are thus called *steepest descent* methods. Although they are guaranteed to reduce χ^2, as they approach a least squares solution $\nabla\chi^2 \rightarrow 0$ and so does $\Delta\mathbf{m}$, and the method becomes very inefficient.

An algorithm to compensate for this behaviour was suggested by Marquardt (1963). Consider a model perturbation given by

$$\Delta\mathbf{m} = [\lambda\mathbf{I} + (\mathbf{WJ})^T\mathbf{WJ}]^{-1}(\mathbf{WJ})^T[\mathbf{W}(\mathbf{d} - f(\mathbf{m_0}))]$$

for another scalar, λ. When $\lambda = 0$ it is easy to see that this reduces to the Gauss-Newton step. When λ is large, it reduces to a steepest descent step with $\lambda = 1/\mu$. Referring to our earlier, compact notation, this is the same as increasing the diagonal terms of the curvature matrix by a factor λ:

$$\alpha_{jk} = \alpha_{jk}(1 + \lambda) \qquad \text{for } j = k$$

$$\alpha_{jk} = \alpha_{jk} \qquad \text{for } j \neq k \quad .$$

By adjusting λ to be large far from the solution, and small as we approach the minimum, we have a fairly stable method of determining a model composed of a small number of parameters.

The classic Marquardt algorithm is as follows:

i) Choose a starting $\mathbf{m_0}$ and a fairly small λ (say, 0.001). Compute the χ^2 of the starting model.

ii) Compute $\mathbf{m_1}$ and a new χ^2. If

a) the χ^2 decreased, keep the model, reduce λ by 10, and go to (iii)

b) the χ^2 increased, increase λ by 10 and go to (ii).

iii) If the change in χ^2 or $\Delta\mathbf{m}$ are very small, stop. Otherwise, go to (ii).

The assumption in the classic Marquardt algorithm is that forward model calculations are expensive. If they are not, or if the cost of computing $\mathbf{J}$ is high compared with the forward solution, then a more efficient algorithm would be:

i) Choose a starting $\mathbf{m_0}$ and a fairly small λ (say, 0.001). Compute the χ^2 of the starting model.

ii) Do a line search over λ to find the minimum χ^2 and the associated model.

iii) If the change in χ^2 or $\Delta\mathbf{m}$ are very small, stop. Otherwise, go to (ii).

What about calculating those derivatives in $\mathbf{J}$? One approach is to do this numerically by first differences. Forward or backward differences

$$\frac{d}{dx}f(x) \approx \frac{f(x+h)-f(x)}{h} \approx \frac{f(x)-f(x-h)}{h}$$

have a relative error term that goes as

$$\epsilon \sim \frac{h}{2}|f''(x)|$$

but a central difference

$$\frac{d}{dx}f(x) \approx \frac{f(x+h)-f(x-h)}{2h}$$

has an error term that goes as

$$\epsilon \sim \frac{h^2}{6}|f'''(x)| \quad .$$

How to choose h? The tradeoff is to generate a computationally significant difference in f while not being so large that nonlinearities compromise the result. Usually something like 5% works well enough, but make sure you don't take a percentage of anything that can go through zero.

Alternatively, one can do the derivative calculations analytically, which is usually computationally much faster and more accurate but may take some math and coding. If you do, check the analytical results against the central differences – if your derivatives are wrong, no gradient-based inverse method is going to converge.

The iterative parameterized LS approach may converge to a *local minimum* rather than a *global minimum,* and it may not work (diverge) unless you start reasonably close to a solution. One approach to this is to run lots of inversions using a fairly large random set of starting models, assuming the computation is fast enough.

For nonlinear problems that are truly parameterized the Marquardt method is pretty hard to beat. It also works fairly well for problems where the number of degrees of freedom is large, given by $M - N$ when the M data are truly independent, and the starting model provides a rough fit to the data.